

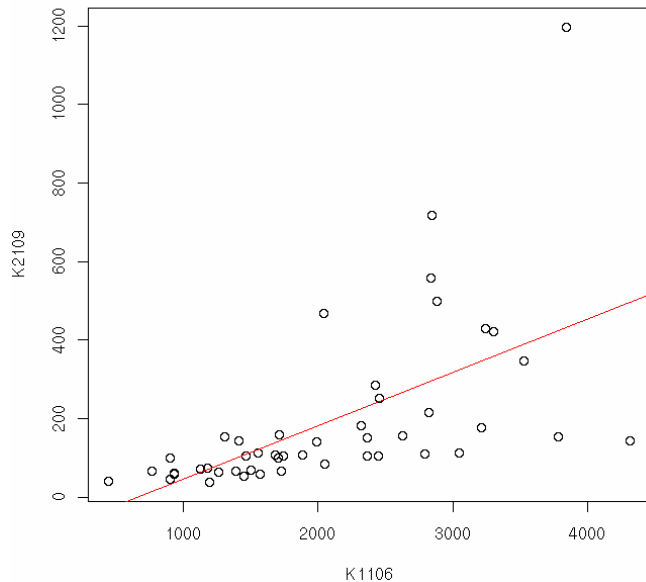
回帰分析のベキ変換

- 単回帰分析: $Y \sim X$
- Y の分布に偏りがあるとき $\log(Y) \sim X$
- 一般化したベキ変換 $Y^\lambda \sim X$
- Y だけでなく X の変換も考慮する

回帰分析 (log変換)

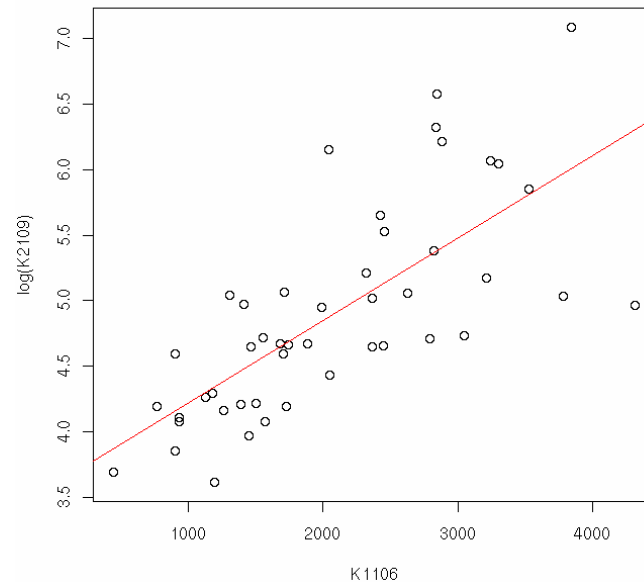
```
> X2000.data <- read.table("2005/X2000data.txt")
> X2000.item <- read.table("2005/X2000item.txt")
> X2000.item[c("K1106", "K2109"), ]
```

	Imi	Tani Zenkoku	Bunrui
K1106	消防ポンプ自動車等現有数	(台)	97368 K. 安全1)消防施設
K2109	火災死傷者数 (人:person)	9037	K. 安全3)火災



変換無し

```
> f1 <- lm(K2109 ~ K1106 , X2000.data)
> plot(K2109 ~ K1106, X2000.data); abline(f1,col=2)
```

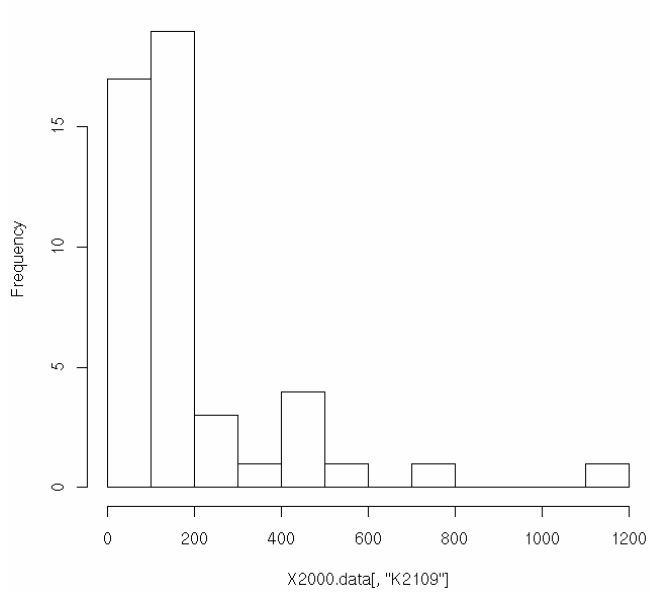


Yを変換

```
> f2 <- lm(log(K2109) ~ K1106 , X2000.data)
> plot(log(K2109) ~ K1106, X2000.data); abline(f2,col=2)
```

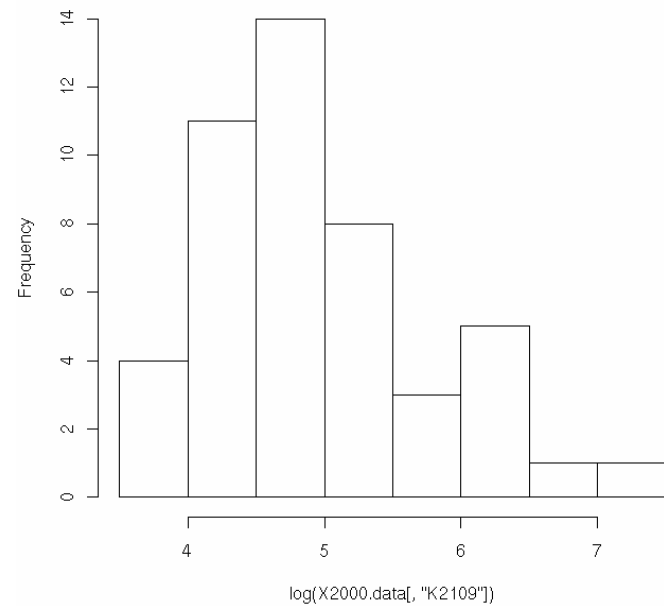
Yのヒストグラム

Histogram of X2000.data[, "K2109"]



変換無し

Histogram of log(X2000.data[, "K2109"])

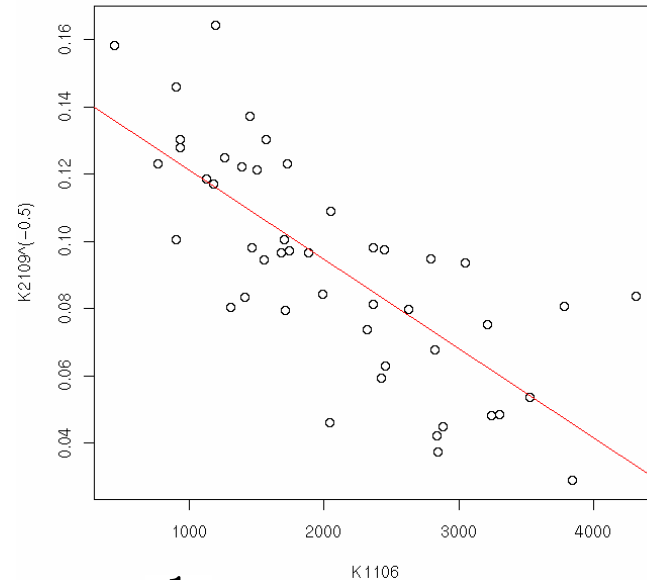
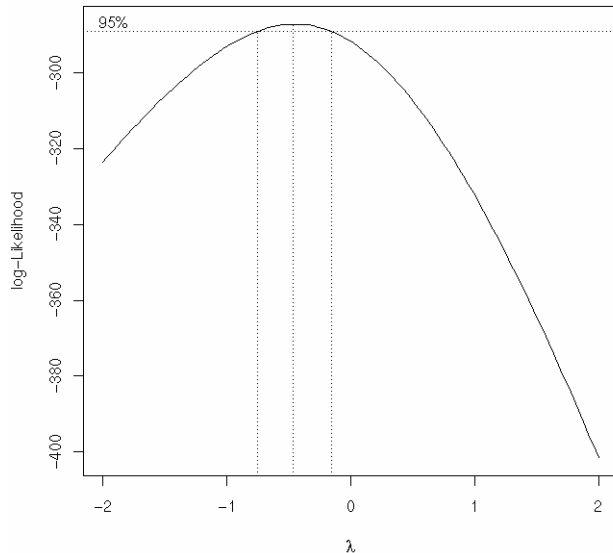


log変換

ベキ変換 (Box-Cox変換)

$$f(y; \lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases}$$

変換の良さ



最適な変換パラメタを探す



$\frac{1}{\sqrt{y}}$ を採用

```
> library(MASS)
> boxcox(K2109 ~ K1106 , data=X2000.data)
> f3 <- lm(K2109(-0.5) ~ K1106 , X2000.data)
> plot(K2109(-0.5) ~ K1106, X2000.data); abline(f3,col=2)
```